Title: Genome-wide association study for vitamin D levels reveals 69 independent loci.

Authors: Despoina Manousaki*[1,2], Ruth Mitchell*[3], Tom Dudding[3,4], Simon Haworth[3,4], Adil Harroud[5], Vince Forgetta[2], Rupal L. Shah[6], Jian'an Luan[6], Claudia Langenberg[6], Nicholas J. Timpson[3] and J. Brent Richards[1,2,7,8,9]

[1] Department of Human Genetics, McGill University, Montreal, Quebec, H3A 1B1, Canada

[2] Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, H3T 1E2, Canada

[3] MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN , United Kingdom

[4] Bristol Dental School, University of Bristol, Bristol, BS8 2BN, United Kingdom

[5] Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, H3A 2B4, Canada

[6]MRC Epidemiology Unit, University of Cambridge, Cambridge, CB2 0SL, United Kingdom

[7] Department of Medicine, McGill University Montreal, Quebec, H3G 1Y6, Canada

[8]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, H3A 1A2, Canada

[9] Department of Twin Research and Genetic Epidemiology, King's College London, London, WC2R 2LS, United Kingdom

* denotes equal contribution

Correspondence:

J. Brent Richards, Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Côte Ste-Catherine Road, Montreal, Quebec H3T 1E2

Tel: 514-340-8222, Fax: 514-340-7502

Email: brent.richards@mcgill.ca

Word count (main text): 3,896

Number of references: 70

Number of figures and tables: 4

31 Number of supplemental tables: 9

Abstract

We aimed to increase our understanding of the genetic determinants of vitamin D levels by undertaking a large-scale genome-wide association study (GWAS) of serum 25 hydroxyvitamin D (25OHD). To do so, we used imputed genotypes from 401,460 white British UK Biobank participants with available 25OHD levels, retaining single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) > 0.1%, and imputation quality score > 0.3. We performed a linear mixed model GWAS on standardized log-transformed 25OHD, adjusting for age, sex, season of measurement and vitamin D supplementation. These results were combined with those from a previous GWAS including 42,274 Europeans. *In silico* functional follow-up of the GWAS results was undertaken to identify enrichment in gene sets, pathways and expression in tissues, and to investigate the partitioned heritability of 25OHD, and its shared heritability with other traits. Using this approach, the SNP heritability of 25OHD was estimated to 16.1%. 138 conditionally independent SNPs were detected (p-value< 6.6×10^{-9}) among which 53 had MAF<5%. Single variant association signals mapped to 69 distinct loci, among which 63 were novel. We identified enrichment in hepatic and lipid metabolism gene pathways, and enriched expression of the 25OHD genes in liver, skin and gastrointestinal tissues. We observed partially shared heritability between 25OHD and socio-economic traits, a feature which may be mediated through time spent outdoors. Therefore, through the largest 25OHD GWAS to date, we identified 63 novel loci, which underline the contribution of genes outside the vitamin D canonical metabolic pathway to the genetic architecture of 25OHD. (250 words)

Introduction

Vitamin D status, as ascertained by 25-hydroxy-vitamin D level (25OHD), is associated with numerous health outcomes[1]. However, it is unclear if lowered 25OHD level plays a causal role in these outcomes and its exact biological mechanisms of action remains unknown[2; 3]. 25OHD is a steroid pro-hormone and a fat-soluble metabolite of cholecalciferol, which is predominately synthesized by exposure to ultra-violet light or obtained from dietary sources including fortified foods, supplements and oily fish. It plays an important role in regulating calcium and phosphorus concentrations and influences cell proliferation, differentiation, apoptosis and has immune modulating effects[4]. Understanding the etiology of low vitamin D levels could have important public health implications by prioritizing individuals who would benefit from supplementation. The body's vitamin D stores are best reflected by serum 25OHD which is influenced not only by diet and exposure to ultra-violet light, but also by age, body mass index, skin color, and numerous factors regulating exposure to ultra-violet B radiation (including season, geographical latitude, skin coverage)[5; 6]. In addition to these environmental factors, classical twin studies show that 50-80% of the variability in the concentration of 25OHD is explained by genetic factors[7; 8] indicating that this is a highly heritable trait.

In recent years, several genome-wide association studies (GWAS) of serum 25OHD have been conducted on participants of Europeans ancestry, with the largest including 79,366 individuals[9]. These studies have identified six common genetic variants (minor allele frequency (MAF) >5%) which are associated with 25OHD level.[9-12] These variants are in loci near genes having an established role in vitamin D synthesis (*DHCR7*/*NADSYN1* [MIM: 602858] (rs12785878) and *CYP2R1* [MIM: 608713] (rs10741657)), transportation (*GC* [MIM: 139200] (rs2282679)) and

degradation (*CYP24A1* [MIM: 126065] (rs17216707)), as well as outside of known vitamin D metabolism pathways, such as *SEC23A* (Sec23 homolog A, coat protein complex II component [MIM: 610511], rs8018720), involved in endoplasmic reticulum (ER)-Golgi protein trafficking, and *AMDHD1* (amidohydrolase domain containing 1, rs10745742) an enzyme involved in the histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolic pathway[9]. Additionally, a low frequency genetic variant (MAF <5%) at *CYP2R1* (rs117913124), with a four-fold larger effect than common variants at that locus was identified through whole-genome sequencing and deep imputation for low-frequency and rare variants[12].

An improved understanding of the genetic determinants of 25OHD has helped re-assess the role of vitamin D in the aetiology of complex diseases, such as musculoskeletal disorders [1], autoimmune disease, such as multiple sclerosis[13-23] and cancer[24] , through methods for causal inference, such as Mendelian randomization (MR) [25; 26]. For example, four separate MR studies have supported a protective effect of vitamin D against multiple sclerosis [12-14; 27], and these results have clinical implications, reflected in recent clinical care guidelines for the use of vitamin D in preventing multiple sclerosis in those at risk, published by the MS Society of Canada[28]. More than 60 MR studies have been published to date utilising genetic variants associated with 25OHD to aid causal effect estimation[29-46]. A deeper understanding of the genetic determinants contributing to variation in circulating vitamin D levels could enable an improved instrumentation of vitamin D in MR studies, allow better genomic prediction of vitamin D levels and provide insights into biological mechanisms.

Although the most recent 25OHD GWAS study on 79,366 Europeans[9] had double the sample size of the previous GWASs, it yielded only two new 25OHD loci (the *SEC23A* and *AMDHD1*), indicating that 25OHD may be a metabolite with a moderately polygenic architecture. In the same study, little of the 25OHD heritability estimated using all common SNPs was explained (SNP heritability of 7.5%), suggesting that much of its heritability remains to be identified. Against this backdrop, we sought to further understand the phenotypic variance explained by genetic variants and investigate the genetic architecture of 25OHD by increasing substantially the GWAS sample size.

We hypothesized that we could identify new genes encoding enzymes, or carrier proteins affecting the levels of this metabolite, unveiling a more polygenic architecture. We therefore undertook a GWAS of serum 25OHD levels in 401,460 White British individuals from UK Biobank and combined results of this GWAS in a meta-analysis with results from a previous GWAS study including up to 42,274 Europeans. Using this approach, we validated previously described 25OHD loci and identified novel genetic determinants of vitamin D. To gain further insight into the genetic control of the vitamin D metabolic pathway, we looked for overlap of our findings with those of the an unpublished GWAS on 1,25-dihydroxyvitamin D, the active form of vitamin D, which is downstream from 25OHD in the vitamin D metabolic pathway (**Figure 1**). We assessed the identified lead 25OHD variants for interaction with season of 25OHD measurement. Finally, we undertook an *in silico* functional follow-up of our GWAS findings, to identify enrichments in gene sets, pathways, and expression in tissues, and explore the partitioned heritability of 25OHD and its shared genetic architecture with other GWAS traits.

Material and Methods

Phenotypes

Between 2006 and 2010 approximately half a million British adults were recruited by UK Biobank[47]. Participants provided biological samples, physical measurements, and answered questionnaires relating to general health and lifestyle. Ethical approval was granted by the Northwest Multi-Centre Research Ethics Committee, and informed consent was obtained from all participants prior to participation.

Data on 25OHD level (in nmol/L) measured using the Diasorin assay were available from 465,415 samples, representing 449,978 UK Biobank participants. Measurements were performed at baseline (2006-2010), and/or the first follow-up visit (2012-2013). In the present study, we used baseline 25OHD measurements from 401,460 individuals from the White British subset of UK Biobank, as defined below. To account for vitamin D supplement use, we adjusted 25OHD levels by subtracting 21.2 nmol/L from the 25OHD measurement in 24,874 vitamin D supplement users, representing 6% of our study cohort (see **Supplemental Material and Methods** for definition of vitamin D supplementation). We used 21.2 nmol/L because it is the mean increase in 25OHD levels conferred by taking daily 400IU of cholecalciferol, the amount of vitamin D most often found in vitamin D supplements[48]. In 3,057 participants treated with vitamin D supplements, 25OHD levels were lower than 10nmol/L (the detection threshold for Diasorin assay) after subtraction, and thus they were set to 10nmol/L. 25OHD levels were then log transformed and standardized to a mean of 0 and standard deviation of 1 (because of skewness in the distribution of 25OHD levels, and to allow comparison with previous 25OHD GWAS). Distribution of the 25OHD levels appears in **Figure S1.**

GWAS

After stringent quality control, the UK Biobank genotypes, imputed to the combined Haplotype Reference Consortium (HRC)[49] and UK10K haplotype resource panel, provided 20,370,874 genetic variants from the autosomes and the X chromosome to test for their association with 25OHD levels. This quality control removed low quality genetic variants, by retaining only SNPs with a minor allele frequency (MAF) > 0.1%, imputation quality score of >0.3 and Hardy–Weinberg $P > 1 \times 10^{-6}$. For details on genotyping and imputation in UK Biobank see the **Supplemental Material and Methods.**

To minimize bias from population stratification, an issue which is particularly relevant in the search for rare genetic variants associated with traits and disease[50], analysis was restricted to individuals of White British ancestry, which comprises the largest single ancestral group represented in the UK Biobank. It is important to distinguish between the self-identified "White British" in UK Biobank, and the White British subset used in our analysis, where the latter was defined using a principal component analysis. Specifically, we previously defined this White British subset using high-quality genotypes, employing FlashPCA [51] and linkage-disequilibrium-pruned HapMap3 SNPs (MAF > 1%, minor allele count > 5, Hardy-Weinberg Equilibrium $P > 1x\ 10^{-6}$), which were projected onto previously computed principal components using the same SNPs set from 1000 Genomes Phase 3 dataset (N=2,504)[52]. Henceforth, whenever the term "White British" appears in this paper, it refers to the White British subset defined as above. Details on this analysis are provided in the **Supplemental Material and Methods.** Descriptive statistics of this White British subset of UK Biobank are detailed in **Table S1.**

We then tested the additive allelic effects of SNPs on 25OHD levels, using a linear mixed-model in the BOLT-LMM software[53]. The model-fitting was performed on hard-called genotypes from 488,377 participants consisting of 803,113 SNPs. Age, sex, season of 25OHD measurement (as a categorical variable; 1 for winter [January to March];2 for spring [April to June];3 for summer [July to September], and 4 for fall [Oct to Dec]), genotype batch, genotype array, and assessment center (as a proxy for latitude) were included as covariates in the BOLT-LMM. We have previously estimated that 6.6×10^{-9} is an appropriate p-value threshold for genome wide significance for analyzing data from the UK Biobank using the above criteria, accounting for multiple testing[52].

Meta-analysis

We compared the results of the GWAS on UK Biobank to those of a previous 25OHD GWAS published by our group (n=42,274 samples of European ancestry)[12], by performing Pearson correlation of the betas of all variants with p-values $<1 \times 10^{-6}$ in both GWAS using the 'cor.test' function in R. We then combined the summary level results of the two GWAS in an inverse variance weighted fixed effects meta-analysis, using the GWAMA[54] software. Of note, in both GWAS, 25OHD levels were first log-transformed and then standardized to a mean of 0 and a standard deviation of 1. This approach allowed the inverse variance weighted meta-analysis of the results. 25OHD levels in both GWAS were adjusted for age, sex, genotyping center, and season of measurement. In the earlier GWAS[12], 25OHD levels were adjusted for BMI. Since BMI is a heritable trait, we elected not to adjust for it in the UK Biobank GWAS, to avoid introducing collider bias. Also, in the present GWAS on UK Biobank, 25OHD measures were adjusted for vitamin D supplementation, since this information was available for all participants, contrarily to

194 the earlier 25OHD GWAS.

195

196 ***Approximate conditional association analysis***

197 To identify conditionally independent SNPs from this meta-analysis, we used GCTA-COJO
198 version 1.91.1[55; 56], which conditions upon the lead SNP per locus by approximating the genotype-
199 phenotype covariance with correlation matrices and summary statistics (**Supplemental Material**
200 **and Methods**). Variants with high collinearity (multiple regression $R^2 > 0.9$) were excluded, and
201 those situated more than 20,000 pairs away were assumed to be independent. A reference sample
202 of 50,000 unrelated white British individuals randomly selected from the UK Biobank was created
203 for a previous GWAS[52], and was used to model patterns of linkage disequilibrium (LD) between
204 variants. We retained as conditionally independent variants those reaching a genome-wide
205 significant p-value pre- and post-conditioning, and with at least one genome-wide significant
206 satellite SNP within 250,000 pairs. These variants were then positionally and functionally
207 annotated to the physically closest gene using the hg19 gene range list, and the Variant Effect
208 Predictor[57] as implemented in PhenoScanner v2.[58]

209

210 **Estimation of variance explained by significant variants and SNP heritability**

211 We estimated the proportion of 25OHD phenotypic variance tagged by all SNPs on the genotyping
212 array (that is, the SNP heritability) using BOLT-REML function [53] in the UK Biobank GWAS. To
213 estimate the variance explained by independent genome-wide significant SNPs (that is, all the
214 genome-wide significant conditionally independent lead SNPs), we summed the variance
215 explained per independent SNP using the formula: variance explained $\approx 2\beta^2 f(1-f)$, where β and

f denote the effect estimate and the effect allele frequency of the allele on a standardized phenotype, respectively[59].

Interaction analysis with season

25OHD levels are affected by the season of their measurement, which is a proxy for exposure to UVB. To assess if there is an effect modification of the 25OHD SNPs by season, we undertook an interaction analysis of our conditionally independent lead SNPs with season of 25OHD assessment in UK Biobank. First, we visually inspected the mean 25OHD concentrations per season (**Figure S2**), and we selected two discrete seasons in order to optimize the comparisons between seasons with higher and lower mean 25OHD levels ("winter"-individuals assessed Jan-Mar (N=98,674), and "summer"-individuals assessed Jul-Sep (N=95,135). Individuals with vitamin D levels assessed in spring (Apr-Jun) and fall (Oct-Dec) were not included in these analyses. Linear regression was conducted under an additive genetic model. The following variables and co-variables were included in the model: standardized log-transformed serum 25OHD adjusted for vitamin D supplementation as the dependent variable; SNP genotype (coded as 0, 1 or 2) as an independent variable; SNP (genotype)* season of 25OHD measurement (coded as a binary variable: 0 for winter and 1 for summer) as an interaction term; age, sex, season of 25OHD measurement as covariates. P-values below a Bonferroni-corrected threshold (0.05/number of COJO-independent SNPs tested for interaction) for the interaction term implied a significant interaction between season and the tested SNP.

Assessment of inflationary bias in GWAS results

By estimating the lambda GC and the LD score regression (LDSR) intercept, BOLT-LMM software estimated the amount of genomic inflation present in the data that was due to residual population stratification, cryptic relatedness, and other latent sources of bias in the UK Biobank GWAS. We used the lambda GC from GWAMA to estimate the genomic inflation in the meta-analysis of the UK Biobank GWAS and compared this with the previous GWAS meta-analysis[12].

***In-silico* functional follow-up**

Functional follow-up of the meta-analysis summary statistics was performed using Complex Trait Genomic-Virtual Lab[60] web application, which implements a variety of follow-up methods for GWAS summary statistics output from the COJO analysis (**Supplemental Material and Methods**). In brief, association between predicted gene transcription and 25OHD was estimated using S-MultiXcan[61] in the MetaXcan package with the default options implemented. Association statistics for the 48 tissues were combined accounting for correlation between tissues to give transcript-level results, and a Bonferroni correction was applied to account for the number of gene transcripts tested. Gene prioritisation, gene set and tissue enrichment analysis were performed using DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits)[62; 63] to identify likely causal genes at associated loci, highlight gene pathways which are over-represented by associated loci in the single variant results and test whether expression of these genes is enriched in specific tissue types. Genetic correlation between 25OHD and a range of other traits available as publicly available GWAS summary statistics was examined using bivariate LDSR[64] implemented in the LD Hub platform[65]. Finally, partitioned heritability by functional annotation with 53 overlapping categories was performed using stratified LDSR using the baseline model from 1000 Genomes phase 3 data (baselineLD_v2.2, February 2019)[64; 66]. Cell specific heritability

was examined using the --h2-cts flag in LDSR and the multi-tissue gene expression file ("Multi_tissue_gene_expr" containing both GTEx data and Franke lab dataset of microarray gene expression)[65]. These final two analyses were restricted to common variants present in HapMap3 (approximately 1,500,000 SNPs), excluding those within the HLA region defined as Chr6: 25000000 to 34000000 bases inclusive.

GWAS on 1,25-dihydroxyvitamin D

Study participants, genotyping and imputation

The Ely Study, established in 1990, is a prospective study of the aetiology of type 2 diabetes and has been described in detail elsewhere. We studied Ely participants with measures of 1,25-dihydroxyvitamin D to estimate genetic effects the active form of vitamin D[67; 68]. Briefly, Ely comprises individuals of European ancestry aged 40-69 years, registered at a single medical practice in Ely, Cambridgeshire, UK and evaluated in 3 phases. All participants of the Ely Study gave their written informed consent and the study was approved by the local ethics committee. Participants at Phase 3 were genotyped using the HumanCoreExome-24 and InfiniumCoreExome arrays. Details of the genotype quality control appear in **Supplemental Material and Methods.** A total of 1,591 samples and 546,486 variants met the quality control criteria. Imputation was performed using the Sanger Imputation Server (pre-phase with EAGLE2 and impute with PBWT pipeline), and the HRC 1.1 reference panel[49]. Additional variants not captured by the HRC reference panel were imputed using a combined UK10K and 1000 Genomes Phase 3 reference panel resulting in data available for >14 million variants.

1,25-dihydroxyvitamin D phenotype and look-up for the 25OHD conditionally independent SNPs

Phase 1 1,25-dihydroxyvitamin D levels and genetic data were available for 748 Ely participants. Levels of 1,25-dihydroxyvitamin D were natural log transformed before regressing with the inclusion of age, sex, body mass index and season as covariates. Residuals from the regression were standardised and used as the final 1,25-dihydroxyvitamin D phenotype. Genetic association analysis was performed for the conditionally independent variants from the 25OHD GWAS meta-analysis using SNPTEST v2.5.4-beta3[69]. Bonferroni adjustment was applied to association test p-values such that variants with GWAS p-values $<4.10 \times 10^{-4}$ (0.05/122) were considered to meet the corrected significance threshold.

Results

GWAS for 25OHD levels

The GWAS in UK Biobank included 401,460 participants and 20,370,874 variants. The genomic control lambda in BOLT-LMM was 1.23, and the LDSR intercept was 1.06 (**Figure S3**). We found a strong correlation between the effect sizes of the UK Biobank GWAS with our previous GWAS meta-analysis[12]. Specifically, we compared the betas of 20,787 SNPs achieving p-values $< 1 \times 10^{-6}$ in both GWAS (minimum MAF 0.3%) and found a coefficient of correlation (r) of 0.88 (**Figure S4**). We then performed a meta-analysis of the two GWAS on a total of 16,668,957 SNPs (**Figure 2**). The lambda GC of the meta-analysis was 1.23. Using approximate conditional analysis as implemented by GCTA–COJO, we observed 138 conditionally independent signals (pre- and post-conditioning p-value$< 6.6 \times 10^{-9}$), mapping to 69 loci (a locus was defined as 1 Mb region around the SNP reaching the lowest p-value), 63 of which were not reported in previous 25OHD GWAS

(**Table 1 and Table S2**). Of these conditionally independent SNPs, 53 (38%) had MAF<5%, and 85 (62%) were common (MAF≥5%). The 53 SNPs with MAF <5% conferred an average absolute effect of 0.23 standard deviations on standardized log transformed 25OHD levels per effect allele, compared to 0.03 standard deviations of the 85 SNPs with MAF≥5% (**Figure S5**).

The total variance explained by the 138 conditionally independent genome-wide significant vitamin D SNPs was 4.9%. When partitioning the variance explained by these lead SNPs into two MAF categories, we found that low-frequency and rare variants explained 1.8% of the variance in 25OHD levels, whereas common variants explained 3.1% of the variance, respectively. The SNP heritability from all SNPs, independent of GWAS p-value, as estimated by BOLT-LMM on 805,426 hard called variants in UK Biobank was 16.1%, indicating that genome-wide significant independent variants capture less than a third of the variance explained in 25OHD levels by all directly genotyped markers.

Look-up of the 25OHD GWAS variants in the 1,25-dihydroxyvitamin D GWAS

We tested 122 out of the 138 conditionally independent variants from the 25OHD GWAS for genetic association with 1,25-dihydroxyvitamin D. The 16 variants that were not tested were not available in the Ely dataset, either because they were not reliably captured through imputation, or had low MAF (<0.001), and no suitable proxy variant could be identified. Among the 122 conditionally independent variants tested in Ely for association with 1,25-dihydroxyvitamin D, only one rs6127099 in the *CYP24A1* locus on chromosome 20 reached the multiple testing corrected threshold for significance (20:52731402:T_A; β=0.231; p=2.5 x 10^{-4})(**Table 1 and**

Table S2). Finally, among the 122 SNPs, 74 SNPs had a consistent direction of effect on 25OHD and on 1,25-dihydroxyvitamin D levels.

Interaction analysis with season

To investigate the hypothesis that the effect of some of the 25OHD variants is modified by season of measurement, we tested the presence of interaction of the 138 conditionally independent variants with season in 193,809 White British participants, whose 25OHD levels were assessed in summer or in winter. We found significant interaction with season in 11 independent SNPs in the *CYP2R1* locus on chromosome 11, and in a single variant in the *SEC23A* locus on chromosome 14 (all p-values below the Bonferroni-corrected threshold of 3.6 x 10^{-4}) (**Table 1 and Table S2**). The strongest interaction was found for rs117913124 (p-value for interaction 1.5 x 10^{-55}), a previously described low frequency variant in *CYP2R1* with large effect on 25OHD levels (absolute GWAS beta per allele of 0.35 units in standardized log-transformed 25OHD). For all 12 SNPs achieving significant interaction p-values, the direction of the beta for the interaction term genotype*season summer was in the same direction as the direction of the beta on 25OHD levels, meaning that the vitamin D lowering effect of these SNPs "blunts" the expected increase in 25OHD in summer.

***In silico* functional follow-up**

Gene prioritisation and enrichment analyses

Gene prioritisation analysis suggested 70 genes with FDR<5% which might plausibly underlie the distribution of association statistics seen in the single variant results. At many loci, genes within the vitamin D metabolism pathway were suggested as plausible candidates. For example, DEPICT

prioritized *DHCR7* at the lead associated chr11:70313961-71239227 locus and *GC* at chr4:72607410-72669758 locus. Interestingly, *ADH6* [MIM:103735] was a plausible candidate at locus chr4:99916771-100274184 suggesting this locus may have pleiotropic effects on vitamin D and alcohol metabolism (**Table S3**).

Gene set enrichment analysis identified enrichment in 418 pre-defined gene sets with a false discovery rate (FDR) < 5%. The strongest statistical evidence for enrichment was in the following gene sets: the alpha-2-HS Glycoprotein (AHSG), a negatively-charged serum glycoprotein that is synthesized by hepatocytes involved in several processes, including endocytosis, brain development, and the formation of bone tissue ($p=4.18x10^{-7}$); the reactome gene set for "metabolism of lipids and lipoprotein" ($p=7.91x10^{-7}$); several genes involved in immune pathways and therefore expressed in the blood such as 'Elastase, Neutrophil Expressed (ELANE)' ($p=8.43x10^{-7}$); the 'Serum albumin (ALB)' ($p=1.19x10^{-6}$), 'Acidic form of complement factor 4 (C4A)' ($p=1.51x10^{-6}$) and 'ENSG00000211949' gene sets, belonging to the immunoglobulin (Ig) heavy chain locus ($p=1.51x10^{-6}$); biosynethic pathways such as "GO:0044283, small molecule biosynthetic process, $p=1.89x10^{-6}$", "GO:0016053, organic acid biosynthetic process, $p=2.29x10^{-6}$; GO:0046394" and "carboxylic acid biosynthetic process, $p=2.29x10^{-6}$"; and finally liver associated pathways including "MP:0000599, enlarged liver, $p=1.33x10^{-6}$", "GO:0001889, liver development, p=3.35x $x10^{-6}$" and "GO:0061008, hepaticobiliary system development, $p=4.15x10^{-6}$" (**Table S4**). Finally, expression of 25OHD genes was enriched in 17 cell types with an FDR < 5%, including cell lines representing the liver (hepatocytes, $p=1.63x10^{-6}$) and skin (keratinocytes, $p=7.73x10^{-3}$). The tissue-specific analysis found greatest evidence for enrichment in the liver ($p=1.34x10^{-6}$) and the gastrointestinal tract ($p=2.22x10^{-3}$) (**Table S5**), which is in accordance with

the fact that 25OHD is hydroxylated in the liver[70], but also conjugates with glucuronide[71] and sulfate[72] to get excreted in the bile and then gets reabsorbed by the enterohepatic circulation. Collectively, these findings suggest that detectable serum 25OHD levels are influenced by a range of metabolic processes within known physiological pathways, but also extending beyond the canonical vitamin D metabolic pathway.

Predicted gene transcription levels

After applying a Bonferroni-corrected multiple testing threshold ($p<1.94x10^{-6}$), varying expression levels at 377 gene transcripts were predicted to influence 25OHD, out of a total of 25,816 that were tested. Results for all gene transcripts are shown in **Figure 3**. This indicates that although there are 69 loci associated with vitamin D phenotype, there are potentially 377 gene transcripts across multiple tissues whose expression may influence vitamin D. The lead associated genetic transcripts using S-MulTiXcan[61] were consistent with the lead association signals in the single variant results, for example identifying association at *NADSYN1* [MIM:608285](Z-test $p<1.81x10^{-309}$); *DHCR7* (Z-test $p<1.15x10^{-245}$); *GC* (Z-test $p<1.81x10^{-309}$); *CYP2R1* (Z-test $p=2.85x10^{-277}$); *UGT1A4* [MIM:606429] (Z-test $p=3.25x10^{-34}$); *PADI1* [MIM: 607934] (Z-test $p=3.64x10^{-23}$). The S-MulTiXcan[61] method integrates information from multiple tissue-specific predictions improving the statistical power over the single variant method and highlights additional transcripts associated with 25OHD, with the strongest evidence in various forms of *Keratin Associated Protein 5* (*KRTAP5* [MIM:608822]) (Z-test $p<1.81x10^{-309}$), a protein coding gene involved in keratinization and has been identified as a potential read through for *NADSYN1*. This adds further evidence that 25OHD is affected through processes beyond the established vitamin D metabolic pathway. Results are shown in **Table S6.**

Genetic correlation

Genetic correlation results for 25OHD were available for 774 traits from the LD hub catalogue[65], including 517 raw traits from UK Biobank and 257 from other GWAS studies and consortia (**Figure 4**). A total of 101 traits passed a multiple testing corrected Bonferroni p-value threshold of $p<6.46x10^{-5}$. The strongest evidence of negative genetic correlation with 25OHD were 'Time spent using a computer' (r_g=-0.22) and 'Qualifications: College or University degree' (r_g=-0.17); 'Intelligence' (r_g=-0.24). Traits pertaining to exercise ('Duration of vigorous activity' (r_g=0.22) and 'Number of days/week walked 10+ minutes' (r_g=0.18)) had positive genetic correlations with vitamin D. Traits related to body mass index (BMI) including lipids and diabetes, had a negative correlation: 'BMI' (r_g=-0.14); 'Triglycerides' (r_g=-0.25); 'Type 2 Diabetes' (r_g=-0.19). A full list of results can be found in **Table S7**.

Tests for enrichment in functional annotations

Using information from all the SNPs in the 25OHD GWAS summary statistics and modelling LD with the 53 functional categories not specific to any cell type in the baseline model, there was evidence for enrichment in 3 out of the 95 functional annotations tested. These were annotations providing evidence for evolutionary conservation with 2% of variants annotated as highly conserved accounting for 20% of the heritability of vitamin D (9-fold enrichment over baseline, $p=1.48x10^{-5}$) (**Table S8**). There was little evidence from stratified LDSR[66] that vitamin D heritability is enriched in gene sets expressed specifically in given cells or tissue types. However, it is worth noting that the highest LDSR coefficients were seen for genomic regions specifically

expressed in hepatocytes (coefficient = $1.17x10^{-8}$), liver (coefficient = $1.73\ x10^{-8}$) and whole blood (coefficient= $1.16\ x10^{-8}$), corroborating the cell and tissue predicted gene enrichment (**Table S9**).

Discussion

This large-scale GWAS meta-analysis identified 63 novel genetic loci which were associated with 25OHD levels in people of European ancestry and at least doubled the estimate of SNP heritability of 25OHD levels. Our study also replicated the 6 known vitamin D loci (in or near *CYP2R1, DHCR7, GC, CYP24A1, AMDHD1, SEC23A). In silico* follow-up identified enrichment in gene sets and pathways mostly independent from canonical vitamin D synthesis and metabolism pathways. Taken together, these results identify new biological pathways that influence 25OHD levels and demonstrate that this metabolite is moderately polygenic.

The large number of low-frequency and rare variants of large effect among the 138 conditionally independent variants of our GWAS is remarkable and suggests that 25OHD levels have a somewhat distinct genetic architecture when compared to other common traits. Specifically, the average absolute effect on 25OHD of the 53 low-frequency and rare variants was at least 7 times larger than the average effect of the 85 common SNPs, but their contribution to the explained variance of 25OHD was smaller than that of the common SNPs (1.8% vs 3.1%). This is not surprising, given the limited frequency of these variants in the general European population. GWAS with larger sample sizes are needed to further dissect the contribution of rare variants with large effects vs common variants with small effects to the variance of 25OHD levels.

The hypothesis-free approach of GWAS has served to highlight the role of lipid biology in 25OHD levels—a fat-soluble hormone. Specifically, among the 69 identified 25OHD loci, 22 loci are related to serum lipid phenotypes. Examples of these loci are the lipase C *(LIPC* [MIM:151670]*)* on chromosome 15, the low density lipoprotein receptor (*LDLR* [MIM:606945]) and the apolipoprotein C1 (*APOC1* [MIM107710]) on chromosome 19, and the cholesteryl ester transfer protein (*CETP* [MIM:118470]) on chromosome 16. Additionally, our gene enrichment analysis prioritized the metabolism of lipids and lipoprotein gene set, and lipid traits were strongly genetically correlated with 25OHD using LDSR. These findings suggest that 25OHD levels share several of the same biological pathways influencing circulating lipids.

We also found enrichment in loci harboring genes associated with skin keratinization. Among these, an interesting finding was the *FLG* [MIM:135940] on the chromosome 1, which encodes fillagrin, a protein which plays an important role in the skin barrier's function, and deregulation of this function might affect vitamin D in the skin, which is also synthetized in the skin. Another locus related to skin keratinization was the *KRTAP5*, which was prioritized by our *in silico* analyses. However, functional follow-up of these novel loci is required, to characterize the causal genes and/or mechanisms underlying the associations with 25OHD levels. Also, we observed enrichment in loci associated with traits outside the vitamin D pathway, which are not directly linked to 25OHD synthesis and metabolism. We can speculate on the exact mechanism of action of these genes on 25OHD—for instance through their effect on time spend outdoors and consequently exposure to sunlight—but follow-up experiments are necessary to validate these hypotheses.

The results of the interaction analysis with season merit some discussion too. We found evidence for significant interaction with multiple independent common, low-frequency and rare SNPs in the *CYP2R1* locus. *CYP2R1* encodes the enzyme responsible for 25-hydroxylation of vitamin D in the liver [70], a necessary step in the conversion of vitamin D synthetized in the skin after exposure to UVB to 25OHD. Therefore, it is not surprising that individuals heterozygous or homozygous for variants in or near *CYP2R1* show a smaller change in their 25OHD levels as a response to season compared to non-carriers. In other words, we observed that carriers of the effect alleles in this locus have steadily lower 25OHD levels, independently of the season of their measurement. We also observed significant interaction with a common SNP in the *SEC23A* gene, which is involved in endoplasmic reticulum (ER)-Golgi protein trafficking. Although the exact mechanism with which *SEC23A* interacts with season to regulate 25OHD levels remains unknown, it might act as a regulator of the enzymatic activity of CYP2R1, which is located in the endoplasmic reticulum. Functional follow-up experiments are warranted to investigate this hypothesis.

The findings of the look-up of the significant 25OHD SNPs in the 1,25-dihyxroxyvitamin D GWAS provide evidence that the two biomarkers of vitamin D in humans have, to a certain extent, a shared genetic component. This may be expected as both biomarkers share at least the same vitamin D catabolic pathway. However, the small sample size of the 1,25-dihydroxyvitamin D GWAS, the only available GWAS on this trait to date, limits the power for characterization of 1,25-dyhydroxyvitamin D loci. We can therefore speculate that there might be a larger overlap of the genetic architecture of the two biomarkers. 1,25-dihydroxyvitamin D is the active metabolite of vitamin D, and although its levels directly regulate the effects of vitamin D on a cellular level,

it remains understudied because of its short half-life, low concentration in blood[73] and the body's ability to buffer 1,25-dihydroxyvitamin D in deficient individuals by increasing parathyroid hormone. In that aspect, any additional evidence, from larger 1,25-dihydroxyvitamin D GWAS, linking 25OHD levels to those of 1,25-dihydroxyvitamin D in the genetic level will be important, as it will add to our understanding of the vitamin D physiology.

Collectively the results of our analyses suggest that serum levels of 25OHD are in crosstalk with a range of metabolic processes extending within the canonical vitamin D metabolic pathway (skin synthesis, hepatic hydroxylation, sulfonylation, glucuronylation), and beyond (time of computer use, intelligence, educational achievement). Although not specifically tested in the present study, one implication of these findings is that the potential genetic instruments for vitamin D are instrumenting more than the vitamin D pathway, and specifically they also capture variance in traits that relate to environmental confounders that could influence 25OHD levels. Taken together, our findings present a cautionary tale for future MR studies using 25OHD as an exposure, based on this GWAS, since there is a risk of pleiotropic effects for a substantial number of novel 25OHD-related SNPs mapping to genes not directly involved in 25OHD biology.

In summary, we described 63 novel loci which are associated with 25OHD levels in Europeans. Further research is warranted to better characterize the novel genetic variants, replicate these findings in independent European samples, validate them in other ethnic groups and identify ancestry-specific variants, and to better understand the biological pathways influencing 25OHD levels. The genetic instruments for 25OHD identified here should be used with caution in future MR analyses assessing the association between vitamin D and other complex traits and diseases.

Supplemental Data

Supplemental Data include Supplemental Material and Methods, 3 Figures and 8 Tables

Declaration of Interests: The authors declare no competing interests.

Ethical approval: All data sources used in this study (UK Biobank, Ely Study) received approval from respective national ethical committees for medical research and obtained informed consent from all participants. Additional ethical approval was not required for this study.

Acknowledgements: This research has been conducted using the UK Biobank Resource under application number 16009. JBR is supported by the Canadian Institutes of Health Research (CIHR), the Canadian Foundation for Innovation and the Fonds de Recherche du Québec – Santé (FRQS). JBR is funded by a FRQS Clinical Research Scholarship and received research support from the National MS Society and the MS Society of Canada. DM is funded by the JDRF. REM works in programme one within the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (grant number MC_UU_00011/1). SH is funded by a National Institute for Health Research (NIHR) Academic Clinical Fellowship. The Ely study was supported by the Medical Research Council (MC_UU_12015/1) and NHS Research and Development. The funding agencies had no role in the study design, analysis, or interpretation of data; the writing of the manuscript; or in the decision to submit the article for publication.

Web resources

OMIM, http://www.omim.org/

Genomic-Virtual Lab , https://genoma.io

The GWAS summary-level results will become available through GRASP

https://grasp.nhlbi.nih.gov/

References

1. Bouillon, R., Marcocci, C., Carmeliet, G., Bikle, D., White, J.H., Dawson-Hughes, B., Lips, P., Munns, C.F., Lazaretti-Castro, M., Giustina, A., et al. (2018). Skeletal and extra-skeletal actions of vitamin D: Current evidence and outstanding questions. Endocr Rev 40, 1109-1151.
2. Theodoratou, E., Tzoulaki, I., Zgaga, L., and Ioannidis, J.P. (2014). Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ 348, g2035.
3. Autier, P., Mullie, P., Macacu, A., Dragomir, M., Boniol, M., Coppens, K., Pizot, C., and Boniol, M. (2017). Effect of vitamin D supplementation on non-skeletal disorders: a systematic review of meta-analyses and randomised trials. Lancet Diabetes Endocrinol 5, 986-1004.
4. Haroon, M., and Fitzgerald, O. (2012). Vitamin D and its emerging role in immunopathology. Clin Rheumatol 31, 199-202.
5. Lagunova, Z., Porojnicu, A.C., Lindberg, F., Hexeberg, S., and Moan, J. (2009). The dependency of vitamin D status on body mass index, gender, age and season. Anticancer Res 29, 3713-3720.
6. Shea, M.K., Benjamin, E.J., Dupuis, J., Massaro, J.M., Jacques, P.F., D'Agostino, R.B., Sr., Ordovas, J.M., O'Donnell, C.J., Dawson-Hughes, B., Vasan, R.S., et al. (2009). Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr 63, 458-464.
7. Karohl, C., Su, S., Kumari, M., Tangpricha, V., Veledar, E., Vaccarino, V., and Raggi, P. (2010). Heritability and seasonal variability of vitamin D concentrations in male twins. Am J Clin Nutr 92, 1393-1398.
8. Hunter, D., De Lange, M., Snieder, H., MacGregor, A.J., Swaminathan, R., Thakker, R.V., and Spector, T.D. (2001). Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J Bone Miner Res 16, 371-378.
9. Jiang, X., O'Reilly, P.F., Aschard, H., Hsu, Y.H., Richards, J.B., Dupuis, J., Ingelsson, E., Karasik, D., Pilz, S., Berry, D., et al. (2018). Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. Nat Commun 9, 260.
10. Wang, T.J., Zhang, F., Richards, J.B., Kestenbaum, B., van Meurs, J.B., Berry, D., Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., et al. (2010). Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 376, 180-188.
11. Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K.C., McCullough, M.L., Gallicchio, L., Jacobs, E.J., Ascherio, A., Helzlsouer, K., Jacobs, K.B., et al. (2010). Genome-wide association study of circulating vitamin D levels. Hum Mol Genet 19, 2739-2745.
12. Manousaki, D., Dudding, T., Haworth, S., Hsu, Y.H., Liu, C.T., Medina-Gomez, C., Voortman, T., van der Velde, N., Melhus, H., Robinson-Cohen, C., et al. (2017). Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. Am J Hum Genet 101, 227-238.

13. Mokry, L.E., Ross, S., Ahmad, O.S., Forgetta, V., Smith, G.D., Goltzman, D., Leong, A., Greenwood, C.M., Thanassoulis, G., and Richards, J.B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. PLoS Med 12, e1001866.
14. Rhead, B., Baarnhielm, M., Gianfrancesco, M., Mok, A., Shao, X., Quach, H., Shen, L., Schaefer, C., Link, J., Gyllenberg, A., et al. (2016). Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. Neurol Genet 2, e97.
15. Giulietti, A., Gysemans, C., Stoffels, K., van Etten, E., Decallonne, B., Overbergh, L., Bouillon, R., and Mathieu, C. (2004). Vitamin D deficiency in early life accelerates Type 1 diabetes in non-obese diabetic mice. Diabetologia 47, 451-462.
16. Riachy, R., Vandewalle, B., Moerman, E., Belaich, S., Lukowiak, B., Gmyr, V., Muharram, G., Kerr Conte, J., and Pattou, F. (2006). 1,25-Dihydroxyvitamin D3 protects human pancreatic islets against cytokine-induced apoptosis via down-regulation of the Fas receptor. Apoptosis 11, 151-159.
17. (1999). Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus. The EURODIAB Substudy 2 Study Group. Diabetologia 42, 51-54.
18. Hypponen, E., Laara, E., Reunanen, A., Jarvelin, M.R., and Virtanen, S.M. (2001). Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet 358, 1500-1503.
19. Mayer-Davis, E.J., Dabelea, D., Crandell, J.L., Crume, T., D'Agostino, R.B., Jr., Dolan, L., King, I.B., Lawrence, J.M., Norris, J.M., Pihoker, C., et al. (2013). Nutritional factors and preservation of C-peptide in youth with recently diagnosed type 1 diabetes: SEARCH Nutrition Ancillary Study. Diabetes Care 36, 1842-1850.
20. Littorin, B., Blom, P., Scholin, A., Arnqvist, H.J., Blohme, G., Bolinder, J., Ekbom-Schnell, A., Eriksson, J.W., Gudbjornsdottir, S., Nystrom, L., et al. (2006). Lower levels of plasma 25-hydroxyvitamin D among young adults at diagnosis of autoimmune type 1 diabetes compared with control subjects: results from the nationwide Diabetes Incidence Study in Sweden (DISS). Diabetologia 49, 2847-2852.
21. Baumgartl, H.J., Standl, E., Schmidt-Gayk, H., Kolb, H.J., Janka, H.U., and Ziegler, A.G. (1991). Changes of vitamin D3 serum concentrations at the onset of immune-mediated type 1 (insulin-dependent) diabetes mellitus. Diabetes Res 16, 145-148.
22. Bierschenk, L., Alexander, J., Wasserfall, C., Haller, M., Schatz, D., and Atkinson, M. (2009). Vitamin D levels in subjects with and without type 1 diabetes residing in a solar rich environment. Diabetes Care 32, 1977-1979.
23. Pozzilli, P., Manfrini, S., Crino, A., Picardi, A., Leomanni, C., Cherubini, V., Valente, L., Khazrai, M., Visalli, N., and group, I. (2005). Low levels of 25-hydroxyvitamin D3 and 1,25-dihydroxyvitamin D3 in patients with newly diagnosed type 1 diabetes. Horm Metab Res 37, 680-683.
24. Feldman, D., Krishnan, A.V., Swami, S., Giovannucci, E., and Feldman, B.J. (2014). The role of vitamin D in reducing cancer risk and progression. Nat Rev Cancer 14, 342-357.
25. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32, 1-22.
26. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 37, 658-665.

27. Gianfrancesco, M.A., Stridh, P., Rhead, B., Shao, X., Xu, E., Graves, J.S., Chitnis, T., Waldman, A., Lotze, T., Schreiner, T., et al. (2017). Evidence for a causal relationship between low vitamin D, high BMI, and pediatric-onset MS. Neurology 88, 1623-1629.
28. Canada., M.S.S.o. (2019). Vitamin D and Multiple Sclerosis Recommendations 2018. In. (
29. Larsson, S.C., Traylor, M., Markus, H.S., and Michaelsson, K. (2018). Serum Parathyroid Hormone, 25-Hydroxyvitamin D, and Risk of Alzheimer's Disease: A Mendelian Randomization Study. Nutrients 10.
30. He, Y., Timofeeva, M., Farrington, S.M., Vaughan-Shaw, P., Svinti, V., Walker, M., Zgaga, L., Meng, X., Li, X., Spiliopoulou, A., et al. (2018). Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. BMC Med 16, 142.
31. Aspelund, T., Grubler, M.R., Smith, A.V., Gudmundsson, E.F., Keppel, M., Cotch, M.F., Harris, T.B., Jorde, R., Grimnes, G., Joakimsen, R., et al. (2019). Effect of Genetically Low 25-Hydroxyvitamin D on Mortality Risk: Mendelian Randomization Analysis in 3 Large European Cohorts. Nutrients 11.
32. Michaelsson, K., Melhus, H., and Larsson, S.C. (2018). Serum 25-Hydroxyvitamin D Concentrations and Major Depression: A Mendelian Randomization Study. Nutrients 10.
33. Bowman, K., Jones, L., Pilling, L.C., Delgado, J., Kuchel, G.A., Ferrucci, L., Fortinsky, R.H., and Melzer, D. (2019). Vitamin D levels and risk of delirium: A mendelian randomization study in the UK Biobank. Neurology 92, e1387-e1394.
34. Lund-Nielsen, J., Vedel-Krogh, S., Kobylecki, C.J., Brynskov, J., Afzal, S., and Nordestgaard, B.G. (2018). Vitamin D and Inflammatory Bowel Disease: Mendelian Randomization Analyses in the Copenhagen Studies and UK Biobank. J Clin Endocrinol Metab 103, 3267-3277.
35. Sun, J.Y., Zhao, M., Hou, Y., Zhang, C., Oh, J., Sun, Z., and Sun, B.L. (2019). Circulating serum vitamin D levels and total body bone mineral density: A Mendelian randomization study. J Cell Mol Med 23, 2268-2271.
36. Jiang, X., Dimou, N.L., Al-Dabhani, K., Lewis, S.J., Martin, R.M., Haycock, P.C., Gunter, M.J., Key, T.J., Eeles, R.A., Muir, K., et al. (2018). Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study. Int J Epidemiol 48, 1416-1424.
37. Larsson, S.C., Traylor, M., Mishra, A., Howson, J.M.M., Michaelsson, K., Markus, H.S., and Consortium, M.P.o.t.I.S.G. (2018). Serum 25-Hydroxyvitamin D Concentrations and Ischemic Stroke and Its Subtypes. Stroke 49, 2508-2511.
38. Yarmolinsky J, R.C., Lophatananon A, et al. . (November 2018:472696.). Evaluating causal associations between previously reported risk factors and epithelial ovarian cancer: a Mendelian randomization analysis. BioRxiv DOI: doi.org/10.1101/472696.
39. Mai, X.M., Videm, V., Sheehan, N.A., Chen, Y., Langhammer, A., and Sun, Y.Q. (2019). Potential causal associations of serum 25-hydroxyvitamin D with lipids: a Mendelian randomization approach of the HUNT study. Eur J Epidemiol 34, 57-66.
40. Dong, J., Gharahkhani, P., Chow, W.H., Gammon, M.D., Liu, G., Caldas, C., Wu, A.H., Ye, W., Onstad, L., Anderson, L.A., et al. (2019). No Association Between Vitamin D Status and Risk of Barrett's Esophagus or Esophageal Adenocarcinoma: A Mendelian Randomization Study. Clin Gastroenterol Hepatol 17,2227-2235.

41. Tan, V.Y., Biernacka, K.M., Dudding, T., Bonilla, C., Gilbert, R., Kaplan, R.C., Qibin, Q., Teumer, A., Martin, R.M., Perks, C.M., et al. (2018). Reassessing the Association between Circulating Vitamin D and IGFBP-3: Observational and Mendelian Randomization Estimates from Independent Sources. Cancer Epidemiol Biomarkers Prev 27, 1462-1471.
42. Havdahl, A., Mitchell, R., Paternoster, L., and Davey Smith, G. (2019). Investigating causality in the association between vitamin D status and self-reported tiredness. Sci Rep 9, 2880.
43. Yuri Milaneschi, W.J.P., Michel G Nivard, Hamdi Mbarek, Dorret I Boomsma, Brenda WJH Penninx. (2019). A role for vitamin D and omega-3 fatty acids in major depression? An exploration using genomics. Transl Psychiatry 9, 219.
44. Libuda, L., Laabs, B.H., Ludwig, C., Buhlmeier, J., Antel, J., Hinney, A., Naaresh, R., Focker, M., Hebebrand, J., Konig, I.R., et al. (2019). Vitamin D and the Risk of Depression: A Causal Relationship? Findings from a Mendelian Randomization Study. Nutrients 11.
45. Liyanage, U.E., Law, M.H., Melanoma Meta-analysis, C., Barrett, J.H., Iles, M.M., and MacGregor, S. (2019). Is there a causal relationship between vitamin D and melanoma risk? : A Mendelian randomization study. Br J Dermatol 182, 97-103.
46. Xiangrui Meng, X.L., Maria Timofeeva, Yazhou He, Athina, and Spiliopoulou, W.-Q.W., Aliya Gifford, Hongjiang Wu, Timothy Varley, Peter Joshi,Joshua C. Denny, Susan Farrington, Lina Zgaga, Malcolm G. Dunlop, Paul McKeigue, Harry Campbell, Evropi Theodoratou. (2019). Phenome Wide Mendelian Randomization Study of genetically determined Vitamin D on multiple health outcomes using the UK Biobank Study. Int J Epidemiol 48, 1425-1434.
47. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203-209.
48. McKenna, M.J., and Murray, B.F. (2013). Vitamin D dose response is underestimated by Endocrine Society's Clinical Practice Guideline. Endocr Connect 2, 87-95.
49. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y, A.R., H, K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48, 1443-1448.
50. Persyn, E., Redon, R., Bellanger, L., and Dina, C. (2018). The impact of a fine-scale population stratification on rare variant association test results. PLoS One 13, e0207677.
51. Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am J Hum Genet 98, 456-472.
52. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet 51, 258-266.
53. Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., Schizophrenia Working Group of Psychiatric Genomics, C., de Candia, T.R., Lee, S.H., Wray, N.R., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet 47, 1385-1392.
54. Magi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. Bmc Bioinformatics 11.

55. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190-2191.
56. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.
57. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol 17, 122.
58. Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth, A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics 35, 4851-4853.
59. Park, J.H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J., and Chatterjee, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42, 570-575.
60. Cuellar-Partida G, L.M., Kho PF, D'Urso S, Gutierrez-Mondragon LF, Hwang L-D. (2019). Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses. bioRxiv DOI: https://doi.org/10.1101/518027.
61. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet 15, e1007889.
62. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 6, 5890.
63. Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hubenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 48, 510-518.
64. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen, C., Psychiatric Genomics, C., Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control, C., Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-1241.
65. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 50, 621-629.
66. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228-1235.
67. Forouhi, N.G., Luan, J., Hennings, S., and Wareham, N.J. (2007). Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: the Ely study 1990-2000. Diabet Med 24, 200-207.
68. Williams, D.R., Wareham, N.J., Brown, D.C., Byrne, C.D., Clark, P.M., Cox, B.D., Cox, L.J., Day, N.E., Hales, C.N., Palmer, C.R., et al. (1995). Undiagnosed glucose intolerance in the community: the Isle of Ely Diabetes Project. Diabet Med 12, 30-35.
69. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat Rev Genet 11, 499-511.

754 70. Cheng, J.B., Levine, M.A., Bell, N.H., Mangelsdorf, D.J., and Russell, D.W. (2004). Genetic
755 evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proc Natl
756 Acad Sci U S A 101, 7711-7715.
757 71. Wang, X., Wang, H., Shen, B., Overholser, B.R., Cooper, B.R., Lu, Y., Tang, H., Zhou, C., Sun,
758 X., Zhong, L., et al. (2016). 1-Alpha, 25-dihydroxyvitamin D3 alters the pharmacokinetics
759 of mycophenolic acid in renal transplant recipients by regulating two extrahepatic UDP-
760 glucuronosyltransferases 1A8 and 1A10. Transl Res 178, 54-62 e56.
761 72. Wong, T., Wang, Z., Chapron, B.D., Suzuki, M., Claw, K.G., Gao, C., Foti, R.S., Prasad, B.,
762 Chapron, A., Calamia, J., et al. (2018). Polymorphic Human Sulfotransferase 2A1
763 Mediates the Formation of 25-Hydroxyvitamin D3-3-O-Sulfate, a Major Circulating
764 Vitamin D Metabolite in Humans. Drug Metab Dispos 46, 367-379.
765 73. Zittermann, A., Schleithoff, S.S., Frisch, S., Gotting, C., Kuhn, J., Koertke, H., Kleesiek, K.,
766 Tenderich, G., and Koerfer, R. (2009). Circulating calcitriol concentrations and total
767 mortality. Clin Chem 55, 1163-1170.
768

769

Figure 1. The vitamin D metabolic pathway

Figure 2. Genome-wide association of 25OHD graphed by chromosome positions and −log10 P-value (Manhattan plot), and quantile-quantile plot of the GWAS meta-analysis (QQ-plot) on 443,374 European individuals. A Manhattan plot: The P-values were obtained from the fixed-effects inverse variance weighted meta-analysis. Horizontal red dash line represents the thresholds of $P = 6.6\times10^{-9}$ for genome-wide significance. Known loci were colored coded as blue diamonds, novel rare loci were color coded as red diamonds, and novel common loci were color coded as white diamonds. B QQ-plot: The Y axis (observed −log10 P-values) is truncated at 310; the X axis shows the expected −log10 P-values. Each SNP is plotted as a blue dot, and the dash red line indicates null hypothesis of no true association. Deviation from the expected P-value distribution is evident only in the tail area, with a lambda of 1.23.

Figure 3. Effect of predicted increased transcription of all genes on circulating vitamin D. Each dot represents the effect of increased transcription (averaged across all tissue-specific predictions using S-MultiXcan) on 25OHD.

Figure 4. Genetic correlation between 25OHD levels and GWAS traits available within LD hub. Each dot represented the Rg between 25OHD and an individual trait. The red dashed line represents the Bonferroni-corrected multiple testing threshold at the 5% level.

Table 1. Association results for 138 conditionally independent SNPs that reach genome-wide significance in the GWAS meta-analysis for 25OHD, and the 1,25 dihydroxyvitamin D GWAS.

CONDITIONALLY INDEPENDENT 25OHD ASSOCIATED VARIANTS				META-ANALYSIS (N=443,734)				GCTA-COJO			Ely 1,25-dihydroxyvitamin D GWAS (N=748)			GxE with season		LOCUS ANNOTATION	
RSID	CHR	BP	EA	EAF	MAF	β	P	β.J	P.J	VAR.J	EAF	β	P	β genotype*season Summer	P	A.GENE	FUNCTION
rs6698680	1	2329661	G	0.46	0.46	-0.01	8.99E-10	-0.01	7.47E-10	0.0001	0.46	-0.06	0.27	-0.018	0.0024	RER1	*intron*
rs3750296	1	17559656	C	0.34	0.34	-0.02	2.09E-24	-0.02	3.04E-24	0.0002	0.34	0.00	0.94	-0.004	0.5694	*PADI1*	*intron*
rs7519574	1	34726552	A	0.18	0.18	0.02	2.09E-11	0.02	4.03E-11	0.0001	0.17	-0.08	0.28	-0.013	0.0958	*RP4-657M3.2*	*intergenic*
rs56044892	1	41830086	T	0.21	0.21	0.02	2.85E-10	0.02	3.13E-10	0.0001	0.21	-0.01	0.92	0.004	0.5716	*FOXO6*	*intron*
rs2934744	1	63048045	A	0.64	0.36	-0.02	3.96E-26	-0.02	4.13E-26	0.0002	0.63	-0.11	0.04	0.000	0.9646	*DOCK7*	*intron*
rs7528419	1	109817192	G	0.22	0.22	0.02	2.41E-16	0.02	2.43E-16	0.0001	0.23	-0.16	0.01	-0.014	0.0461	*CELSR2*	*3_prime_UTR*
rs3768013	1	150815411	A	0.37	0.37	-0.01	1.37E-13	-0.01	3.86E-09	0.0001	0.38	-0.06	0.30	0.008	0.2182	*ARNT*	*intron*
rs115045402	1	152029548	A	0.03	0.03	0.11	3.05E-55	0.07	1.58E-19	0.0003	0.02	0.20	0.36	-0.006	0.8271	*FLG*	*intergenic*
rs12123821	1	152179152	T	0.05	0.05	0.07	2.25E-59	0.05	1.28E-24	0.0003	0.04	0.04	0.78	-0.029	0.0391	*FLG*	*intron*
rs201561609	1	152187902	T	0.99	0.01	-0.13	6.99E-28	-0.10	6.63E-16	0.0002	0.98	0.02	0.93	-0.027	0.5813	*FLG*	*missence*
rs185433896	1	152249021	A	0.99	0.01	-0.25	1.50E-38	-0.21	7.24E-28	0.0006	1.00	-0.24	0.69	-0.128	0.2258	*FLG*	*intron*
rs189918701	1	152254152	G	1.00	0.00	-0.24	2.47E-16	-0.18	3.29E-10	0.0002	Not available in Ely datasets			-0.077	0.4942	*FLG*	*intron*
rs375984409	1	152255772	G	0.99	0.01	-0.23	3.22E-38	-0.19	1.53E-25	0.0006	Not available in Ely datasets			0.055	0.4834	*FLG*	*intron*
rs144613541	1	152270875	G	0.29	0.29	0.02	6.49E-12	0.02	1.52E-12	0.0001	0.28	0.12	0.05	0.008	0.2585	*FLG*	*downstream*
rs150597413	1	152277622	T	0.00	0.00	0.10	6.18E-11	0.11	1.56E-12	0.0001	0.00	0.13	0.82	0.034	0.4693	*FLG*	
rs138726443	1	152280023	A	0.00	0.00	0.11	8.81E-15	0.12	1.36E-17	0.0001	0.01	0.48	0.25	-0.084	0.0557	*FLG*	
rs61816761	1	152285861	A	0.02	0.02	0.13	8.57E-74	0.11	5.39E-54	0.0005	0.02	-0.01	0.96	-0.021	0.3844	*FLG*	*stop_lost*
rs576242124	1	152390763	A	0.01	0.01	0.11	3.08E-15	0.09	2.59E-10	0.0002	0.01	0.29	0.42	-0.061	0.2780	*FLG*	*upstream*
rs184958517	1	153111312	T	0.99	0.01	-0.13	5.55E-15	-0.11	1.21E-09	0.0002	1.00	0.17	0.79	0.017	0.8560	*FLG*	*downstream*
rs558560635	1	153147997	G	1.00	0.00	-0.27	5.83E-16	-0.24	4.45E-13	0.0003	Not available in Ely datasets			-0.001	0.9953	*FLG*	*intron*
rs11264360	1	155284586	A	0.24	0.24	0.02	3.34E-15	0.02	1.12E-15	0.0001	0.23	-0.09	0.16	-0.011	0.1237	*FDPS*	*indels*
rs867772	1	220972343	G	0.68	0.32	-0.01	3.64E-11	-0.01	3.31E-11	0.0001	0.69	0.00	0.97	0.001	0.8839	*MARC_1*	*intron*
rs10127775	1	230295789	T	0.60	0.40	0.01	3.43E-09	0.01	3.11E-09	0.0001	0.60	0.01	0.87	0.016	0.0074		
rs12997242	2	21381177	A	0.44	0.44	-0.01	2.23E-10	-0.01	2.32E-10	0.0001	0.43	-0.01	0.89	-0.008	0.1958	*TDRD15*	*intergenic*
rs11127048	2	27752463	A	0.62	0.38	0.02	6.41E-19	0.02	6.72E-19	0.0002	0.63	0.03	0.63	0.004	0.4918	*GCKR*	*intergenic*
rs6724965	2	101440151	G	0.17	0.17	-0.02	1.29E-10	-0.02	1.34E-10	0.0001	0.18	0.07	0.31	-0.001	0.9476	*NPAS2*	*intron*
rs7569755	2	118648261	A	0.29	0.29	0.01	8.03E-11	0.01	8.35E-11	0.0001	0.28	-0.03	0.64	0.000	0.9838	*HTR5BP*	*intron*
rs1047891	2	211540507	A	0.32	0.32	-0.01	1.16E-11	-0.01	1.16E-11	0.0001	0.32	-0.01	0.81	-0.004	0.4934	*CPS1*	*missence*
rs2011425	2	234627608	G	0.08	0.08	-0.05	9.66E-38	-0.05	9.93E-38	0.0003	0.06	0.08	0.45	-0.002	0.8714	*UGT1A4*	*missence*
rs7650253	3	49431160	A	0.69	0.31	0.01	1.76E-10	0.01	1.76E-10	0.0001	0.69	0.00	0.99	-0.017	0.0126	*RHOA*	*intron*
rs1972994	3	85631142	T	0.65	0.35	-0.02	7.99E-18	-0.02	8.04E-18	0.0001	0.67	-0.11	0.05	-0.005	0.4647	*CADM2*	*intron*
rs6438900	3	125148287	G	0.26	0.26	0.01	9.59E-10	0.01	1.16E-09	0.0001	0.25	-0.01	0.93	-0.014	0.0391	*MRPL3*	*intergenic*
rs6773343	3	141825598	T	0.72	0.28	0.01	5.20E-09	0.01	6.28E-09	0.0001	0.72	0.02	0.76	0.001	0.8707	*TFDP2*	*intron*
rs78649910	4	3482213	A	0.11	0.11	-0.02	4.32E-09	-0.02	3.41E-09	0.0001	0.12	-0.02	0.79	0.007	0.4484	*DOK7*	*intron*
rs7699711	4	69947596	T	0.45	0.45	-0.03	6.97E-49	-0.03	4.85E-50	0.0004	0.43	0.01	0.88	0.000	0.9588	*UGT2B7*	*intron*
rs529640451	4	72177044	C	1.00	0.00	0.23	2.25E-17	0.17	2.20E-10	0.0002	Not available in Ely datasets			-0.165	0.1887	*GC*	*intergenic*
rs528776789	4	72486140	A	0.99	0.01	0.18	3.67E-31	0.12	2.45E-15	0.0002	0.99	0.06	0.90	0.053	0.4581	*GC*	*intergenic*
rs113938679	4	72488025	A	0.01	0.01	-0.18	5.88E-36	-0.10	2.21E-11	0.0001	0.01	0.20	0.65	0.042	0.4317	*GC*	*intergenic*
rs564377207	4	72488525	G	1.00	0.00	-0.20	1.05E-21	-0.16	2.23E-14	0.0002	1.00	-0.64	0.22	-0.013	0.9058	*GC*	*intergenic*
rs186897112	4	72528565	G	1.00	0.00	0.25	3.79E-13	0.20	3.81E-09	0.0002	Not available in Ely datasets			-0.147	0.3323	*GC*	*intergenic*
rs557657187	4	72539857	G	1.00	0.00	0.37	6.18E-16	0.29	2.19E-10	0.0002	Not available in Ely datasets			-0.274	0.0977	*GC*	*intergenic*

rs145432346	4	72575017	C	0.83	0.17	0.11	6.78E-286	0.03	2.26E-27	0.0003	0.82	0.20	0.01	0.004	0.7215	*GC*	*intergenic*
rs705117	4	72608115	T	0.85	0.15	-0.03	1.71E-36	0.03	1.12E-27	0.0003	0.87	0.06	0.47	0.002	0.7808	*GC*	*intron*
rs11723621	4	72615362	G	0.29	0.29	-0.19	2.903E-1689	-0.16	0	0.0101	0.29	-0.08	0.19	0.011	0.0871	*GC*	*intron*
rs560384646	4	72616618	C	0.02	0.02	-0.19	6.91E-112	-0.09	3.23E-24	0.0004	0.02	-0.54	0.07	0.022	0.4814	*GC*	*indel*
rs200641845	4	72620895	T	0.55	0.45	0.02	6.92E-14	0.02	5.23E-12	0.0001	0.56	-0.16	0.02	-0.022	0.0113	*GC*	*intron*
rs565277381	4	72625772	T	1.00	0.00	0.31	6.62E-11	0.28	3.55E-09	0.0002	Not available in Ely datasets			-0.144	0.4506	*GC*	*intron*
rs3775150	4	72640750	C	0.26	0.26	-0.09	3.90E-295	-0.07	3.46E-109	0.0019	0.27	0.03	0.68	-0.002	0.7781	*GC*	*indel*
rs222026	4	72643760	T	0.87	0.13	-0.05	6.98E-68	-0.05	1.09E-40	0.0006	0.86	-0.05	0.50	0.012	0.2171	*GC*	*intron*
rs190688847	4	72705716	C	1.00	0.00	0.29	1.02E-18	0.25	1.26E-14	0.0003	Not available in Ely datasets			0.002	0.9879	*GC*	*intergenic*
rs184291421	4	72752846	C	0.99	0.01	0.17	1.25E-28	0.09	5.03E-09	0.0001	1.00	-0.38	0.39	-0.064	0.2998	*GC*	*intergenic*
rs188838036	4	72783385	A	1.00	0.00	0.18	3.07E-24	0.12	3.14E-11	0.0001	0.99	0.54	0.20	0.007	0.9173	*GC*	*intergenic*
rs186881826	4	72785743	A	0.22	0.22	0.05	3.64E-77	0.02	1.43E-15	0.0001	0.23	-0.05	0.44	0.000	0.9645	*GC*	*intergenic*
rs186441690	4	72820969	G	1.00	0.00	-0.27	1.96E-18	-0.23	1.79E-14	0.0003	Not available in Ely datasets			0.280	0.0786	*GC*	*intergenic*
rs546541682	4	72864566	T	0.99	0.01	-0.16	2.06E-18	-0.11	3.45E-10	0.0001	0.99	0.37	0.46	0.059	0.5093	*GC*	*intergenic*
rs143106299	4	72920085	T	0.01	0.01	-0.17	1.50E-28	-0.09	4.62E-09	0.0001	0.00	-0.02	0.97	0.126	0.0913	*GC*	*intron*
rs192785674	4	73505826	A	1.00	0.00	0.17	8.14E-11	0.18	3.48E-12	0.0002	Not available in Ely datasets			-0.116	0.4965	*GC*	*intergenic*
rs58073039	4	88287363	G	0.30	0.30	-0.01	2.16E-11	-0.01	2.84E-10	0.0001	0.28	-0.04	0.45	0.015	0.0224	*HSD17B11*	*intron*
rs28364331	4	100201295	G	0.02	0.02	0.06	1.31E-17	0.06	3.06E-18	0.0001	0.02	0.08	0.70	0.050	0.0227	*ADH1A*	*splice_region*
rs1229984	4	100239319	C	0.97	0.03	-0.05	4.85E-13	-0.05	2.43E-13	0.0001	0.97	-0.04	0.84	0.014	0.4919	*ADH1A*	*missence*
rs7718395	5	118652574	G	0.32	0.32	0.01	1.67E-09	0.01	1.68E-09	0.0001	0.31	0.04	0.52	0.006	0.3912	*TNFAIP8*	*intron*
rs3822868	6	131934986	G	0.84	0.16	0.02	1.41E-15	0.02	1.41E-15	0.0001	0.84	0.15	0.02	-0.010	0.2213	*MED23*	*intron*
rs111529171	7	21571932	C	0.22	0.22	-0.02	6.24E-11	-0.02	6.26E-11	0.0001	0.22	0.04	0.50	0.000	0.9982	*DNAH11*	*intergenic*
rs1011468	7	104613791	A	0.48	0.48	-0.01	1.35E-12	-0.01	1.39E-12	0.0001	0.44	-0.12	0.02	0.013	0.0327	*LINC01004*	*intron*
rs1858889	7	107117447	C	0.50	0.50	0.01	3.85E-11	0.01	3.03E-11	0.0001	0.50	-0.02	0.66	-0.010	0.1046	*COG5*	*intron*
rs804280	8	11612698	A	0.58	0.42	0.01	4.43E-11	0.02	9.90E-16	0.0001	0.57	-0.06	0.22	-0.014	0.0207	*GATA4*	*intron*
rs34726834	8	25889606	T	0.25	0.25	0.01	6.65E-10	0.01	3.39E-10	0.0001	0.27	-0.04	0.48	-0.010	0.1456	*EBF2*	*intron*
rs7828742	8	116960729	G	0.60	0.40	-0.02	3.06E-28	-0.02	2.85E-33	0.0003	0.60	-0.03	0.51	0.000	0.9401	*LINC00536*	*downstream*
rs10818769	9	125719923	G	0.86	0.14	-0.02	3.35E-09	-0.02	2.99E-09	0.0001	0.84	0.04	0.59	-0.001	0.9333	*DNAH11*	*intergenic*
rs532436	9	136149830	A	0.18	0.18	-0.02	2.17E-09	-0.02	1.94E-09	0.0001	0.21	0.04	0.55	-0.014	0.0590	*ABO*	*intron*
rs10887718	10	82042624	T	0.53	0.47	-0.01	1.44E-10	-0.01	1.18E-10	0.0001	0.51	0.03	0.54	-0.001	0.8903	*MAT1A*	*intron*
rs538325438	11	13414030	C	1.00	0.00	0.23	6.07E-13	-0.45	4.61E-32	0.0006	Not available in Ely datasets			0.111	0.2836	*CYP2R1*	*intron*
rs373514022	11	13955649	C	1.00	0.00	0.20	4.77E-12	0.21	4.15E-13	0.0002	1.00	-0.89	0.19	-0.067	0.5386	*CYP2R1*	*intergenic*
rs571618690	11	13996822	A	1.00	0.00	0.37	1.90E-31	0.23	1.40E-12	0.0001	Not available in Ely datasets			0.517	2.7E-06	*CYP2R1*	*intron*
rs191379475	11	14075712	G	0.99	0.01	-0.10	1.70E-15	-0.09	1.22E-11	0.0002	0.99	-0.33	0.43	-0.138	0.0417	*CYP2R1*	*intron*
rs561089663	11	14100539	G	1.00	0.00	0.41	4.79E-43	0.20	4.31E-11	0.0001	Not available in Ely datasets			0.078	0.5016	*CYP2R1*	*intron*
rs10832218	11	14181174	C	0.20	0.20	-0.03	7.09E-32	-0.02	3.06E-10	0.0001	0.17	-0.12	0.20	-0.011	0.3662	*CYP2R1*	*intron*
rs117206369	11	14335876	T	1.00	0.00	0.47	1.07E-48	0.23	1.10E-12	0.0002	1.00	-0.39	0.63	0.373	0.0004	*CYP2R1*	*intron*
rs567876843	11	14414139	G	1.00	0.00	0.54	1.83E-180	0.54	3.35E-116	0.0027	0.99	0.34	0.35	0.193	0.0095	*CYP2R1*	*intergenic*
rs148514005	11	14464878	T	0.01	0.01	-0.45	1.37E-184	-0.14	4.99E-15	0.0002	0.01	-0.01	0.97	-0.274	1.3E-06	*CYP2R1*	*downstream*
rs571484036	11	14512559	A	1.00	0.00	-0.22	4.13E-16	-0.25	3.43E-20	0.0002	Not available in Ely datasets			-0.372	1.5E-05	*CYP2R1*	*intron*
rs577185477	11	14612563	C	0.01	0.01	-0.38	1.624E-342	-0.15	7.55E-37	0.0007	0.01	-0.20	0.41	-0.260	1.7E-14	*CYP2R1*	*intron*
rs554808052	11	14636390	C	1.00	0.00	0.35	5.41E-40	0.20	7.88E-13	0.0001	1.00	0.91	0.09	0.438	6.4E-07	*CYP2R1*	*intron*
rs10832289	11	14669496	T	0.41	0.41	-0.07	2.03E-266	-0.09	0	0.0036	0.42	-0.08	0.14	-0.086	1.2E-46	*CYP2R1*	*intron*
rs187443664	11	14768892	T	0.99	0.01	-0.11	3.49E-16	-0.08	1.52E-09	0.0001	0.99	-0.41	0.29	-0.167	0.0102	*CYP2R1*	*intron*
rs188480917	11	14785870	G	0.01	0.01	-0.34	5.00E-275	-0.17	3.21E-37	0.0006	0.01	-0.37	0.12	-0.291	4.2E-21	*CYP2R1*	*intron*
rs534042887	11	14818258	G	1.00	0.00	0.39	2.82E-82	0.19	2.21E-19	0.0002	1.00	0.69	0.22	0.255	0.0006	*CYP2R1*	*intron*
rs532836473	11	14822853	G	1.00	0.00	0.44	4.90E-44	0.27	4.77E-17	0.0002	1.00	-0.16	0.78	0.268	0.0264	*CYP2R1*	*intron*
rs201501563	11	14882470	T	0.12	0.12	-0.07	9.17E-67	-0.03	1.96E-18	0.0003	0.13	-0.07	0.54	-0.112	8.3E-14	*CYP2R1*	
rs117913124	11	14900931	A	0.03	0.03	-0.35	1.653E-775	-0.21	2.94E-107	0.0023	0.04	-0.27	0.04	-0.284	1.5E-55	*CYP2R1*	*synonymous*
rs117576073	11	14912573	T	0.01	0.01	-0.11	1.22E-38	-0.17	1.40E-78	0.0007	0.01	-0.06	0.88	-0.135	3.2E-07	*CYP2R1*	*5_prime_UTR*
rs150585703	11	14951216	G	1.00	0.00	0.48	7.16E-125	0.24	1.56E-27	0.0005	0.99	-0.16	0.72	0.276	7.3E-05	*CYP2R1*	*intron*
rs574992951	11	16580958	C	0.99	0.01	0.09	4.04E-09	0.09	1.69E-09	0.0001	0.99	-0.43	0.22	0.089	0.1615	*PLEKHA7*	*intron*
rs567415847	11	16854631	G	1.00	0.00	0.28	1.03E-14	0.30	1.88E-16	0.0004	Not available in Ely datasets			-0.236	0.1177	*PLEKHA7*	*intron*
rs523583	11	66070146	C	0.47	0.47	0.01	5.58E-10	0.01	6.60E-12	0.0001	0.46	0.02	0.66	-0.002	0.6907	*TMEM151A*	*intergenic*

rs12803256	11	71132868	G	0.77	0.23	0.10	8.599E-407	0.09	1.64E-195	0.0027	0.76	0.06	0.30	-0.010	0.1527	*FLJ42102*	*non_coding_transcript_exon*
rs536006581	11	71135151	G	0.01	0.01	-0.17	8.87E-35	-0.11	5.64E-14	0.0002	0.01	0.62	0.14	0.086	0.0640	*FLJ42102*	*downstream*
rs574615332	11	71144427	A	1.00	0.00	-0.29	1.38E-28	-0.21	5.87E-15	0.0002	Not available in Ely datasets			-0.052	0.6369	*FLJ42102*	*intron*
rs549940584	11	71222408	T	0.01	0.01	0.18	2.31E-72	0.15	1.93E-45	0.0006	0.00	-0.51	0.37	-0.054	0.0992	*FLJ42102*	*intron*
rs200454003	11	71228990	T	0.26	0.26	-0.09	3.68E-256	-0.03	3.49E-21	0.0003	0.29	-0.01	0.93	0.021	0.0118	*FLJ42102*	*intron*
rs10793129	11	75459865	A	0.09	0.09	0.02	1.64E-12	0.03	4.11E-13	0.0001	0.08	0.00	0.99	0.009	0.3882	*RP11-21L23.4*	*intergenic*
rs1149605	11	76485216	C	0.17	0.17	0.02	7.34E-14	0.02	3.36E-15	0.0001	0.18	0.01	0.82	0.025	0.0018	*RP11-21L23.4*	*intergenic*
rs964184	11	116648917	C	0.86	0.14	0.04	5.11E-44	0.04	1.30E-43	0.0004	0.85	0.05	0.53	0.015	0.0853	*ZPR1*	*3_prime_UTR*
rs2847500	11	120114421	A	0.12	0.12	-0.02	7.79E-13	-0.02	1.93E-12	0.0001	0.12	-0.08	0.40	-0.003	0.7323	*ZPR1*	*intron*
rs12317268	12	21352541	G	0.15	0.15	-0.02	9.15E-12	-0.02	9.20E-12	0.0001	0.14	0.13	0.09	-0.007	0.3751	*SLCO1B1*	*intron*
rs9668081	12	38602911	T	0.47	0.47	0.01	5.38E-09	0.01	5.40E-09	0.0001	0.49	0.04	0.44	0.011	0.0601	*FAM166AP9*	*upstream*
rs61937878	12	96371731	T	0.01	0.01	0.12	4.43E-22	0.10	5.63E-17	0.0001	0.01	0.15	0.57	-0.038	0.3136	*HAL*	*missence*
rs10859995	12	96375682	C	0.58	0.42	-0.04	7.03E-89	-0.04	3.03E-91	0.0008	0.58	-0.07	0.18	0.003	0.6502	*HAL*	*intron*
rs8018720	14	39556185	C	0.82	0.18	-0.03	4.04E-36	-0.03	4.10E-36	0.0003	0.84	-0.09	0.20	-0.046	2.6E-09	*SEC23A*	*missence*
rs261291	15	58680178	C	0.36	0.36	-0.02	2.89E-28	-0.02	2.46E-29	0.0002	0.37	-0.01	0.80	0.005	0.4603	*LIPC*	*intron*
rs1800588	15	58723675	T	0.21	0.21	-0.03	2.65E-36	-0.03	3.17E-37	0.0003	0.21	-0.10	0.12	-0.001	0.9433	*LIPC*	*intron*
rs17765311	15	63789952	C	0.34	0.34	-0.02	1.35E-13	-0.02	1.18E-13	0.0001	0.36	0.04	0.47	0.000	0.9895	*AC007950.2*	*downstream*
rs62007299	15	77711719	A	0.71	0.29	-0.01	1.69E-11	-0.01	3.33E-11	0.0001	0.69	0.00	1.00	0.006	0.3478	*PEAK1*	*intron*
rs8063706	16	11909552	T	0.27	0.27	0.01	3.64E-09	0.01	4.27E-09	0.0001	0.29	0.03	0.60	-0.013	0.0442	*BCAR4*	*downstream*
rs77924615	16	20392332	A	0.20	0.20	-0.02	1.46E-10	-0.02	2.28E-10	0.0001	0.20	0.21	0.00	-0.002	0.8408	*PDILT*	*intron*
rs71383766	16	30930233	T	0.42	0.42	0.01	1.15E-09	0.01	1.86E-09	0.0001	0.45	0.03	0.58	-0.013	0.0457	*FBXL19*	*upstream*
rs1800775	16	56995236	A	0.49	0.49	-0.02	1.56E-17	-0.02	1.57E-17	0.0001	0.46	-0.03	0.55	0.000	0.9495	*CETP*	*upstream*
rs2909218	17	66464546	T	0.79	0.21	0.02	2.81E-12	0.02	2.82E-12	0.0001	0.80	0.11	0.10	-0.003	0.6766	*RP11-120M18.2*	*intron*
rs8091117	18	28919794	A	0.07	0.07	-0.02	1.03E-09	-0.02	9.48E-10	0.0001	0.07	0.04	0.69	-0.012	0.3137	*DSG1*	*missence*
rs2037511	18	61366207	A	0.17	0.17	0.02	9.29E-10	0.02	8.35E-10	0.0001	0.16	0.01	0.87	-0.005	0.5735	*SERPINB11*	*intron*
rs57631352	19	4338173	G	0.30	0.30	-0.01	1.48E-09	-0.01	1.50E-09	0.0001	0.31	0.00	0.94	0.010	0.1425	*STAP2*	*intron*
rs73015021	19	11192915	G	0.12	0.12	0.02	1.15E-14	0.02	6.29E-14	0.0001	0.13	0.04	0.59	-0.001	0.9190	*LDLR*	*intergenic*
rs10500209	19	11979164	C	0.28	0.28	-0.01	6.18E-10	-0.01	2.73E-09	0.0001	0.28	-0.08	0.18	0.001	0.8869	*LDLR*	*missence*
rs58542926	19	19379549	T	0.08	0.08	0.03	8.57E-19	0.03	2.63E-19	0.0002	0.07	0.09	0.35	0.006	0.5708	*TM6SF2*	*missence*
rs3814995	19	36342212	T	0.31	0.31	-0.01	2.83E-12	-0.02	1.08E-12	0.0001	0.32	-0.06	0.40	0.006	0.3743	*NPHS1*	*missence*
rs1065853	19	45413233	T	0.08	0.08	0.03	8.32E-14	0.03	2.24E-14	0.0001	0.09	0.01	0.87	-0.008	0.4807	*APOC1*	*upstream*
rs157595	19	45425460	G	0.61	0.39	-0.02	2.95E-14	-0.02	4.25E-15	0.0001	0.62	-0.14	0.01	-0.004	0.5361	*APOC1*	*downstream*
rs112285002	19	48374320	T	0.16	0.16	0.06	1.77E-110	0.06	1.49E-90	0.0008	0.15	0.06	0.36	0.003	0.7114	*SULT2A1*	*3_prime_UTR*
rs62130059	19	48461240	C	0.34	0.34	-0.03	9.25E-34	-0.02	2.64E-12	0.0001	0.32	-0.02	0.78	-0.003	0.7023	*SULT2A1*	*intergenic*
rs10426	19	51517798	A	0.21	0.21	0.03	3.31E-26	0.03	1.59E-26	0.0002	0.20	0.03	0.63	-0.002	0.7403	*KLK10*	*3_prime_UTR*
rs8103262	19	53065814	C	0.31	0.31	0.01	3.18E-09	0.01	6.80E-10	0.0001	0.30	-0.02	0.71	0.005	0.4445	*ZNF808*	*intron*
rs6123359	20	52714706	G	0.11	0.11	0.03	7.74E-24	0.02	7.48E-14	0.0001	0.11	0.02	0.82	0.005	0.6144	*RP13-379L11.3*	*intergenic*
rs6127099	20	52731402	T	0.28	0.28	-0.04	9.30E-62	-0.03	2.22E-32	0.0003	0.29	0.23	0.00	0.012	0.0892	*RP13-379L11.3*	*intergenic*
rs2585442	20	52737123	G	0.25	0.25	0.03	6.87E-49	0.02	3.96E-23	0.0002	0.23	-0.05	0.42	-0.002	0.8239	*RP13-379L11.3*	*intergenic*
rs2762942	20	52788925	A	0.94	0.06	0.05	7.99E-35	0.04	1.69E-23	0.0002	0.94	-0.08	0.52	-0.004	0.7518	*RP13-379L11.3*	*intron*
rs2229742	21	16339172	C	0.10	0.10	-0.03	7.13E-16	-0.03	7.16E-16	0.0001	0.10	0.05	0.59	-0.009	0.3346	*NRIP1*	*missence*
rs2074735	22	31535872	C	0.06	0.06	0.03	6.55E-12	0.03	7.12E-12	0.0001	0.07	-0.05	0.63	-0.023	0.0549	*PLA2G3*	*missence*
rs960596	22	41393520	T	0.34	0.34	0.01	2.23E-09	0.01	2.43E-09	0.0001	0.34	-0.05	0.40	-0.003	0.6047	*SCUBE1*	*intergenic*

Grey and white fonts are used to demarcate variants in the same locus. N: sample size; RSID: reference SNP cluster ID; CHR: chromosome; BP: base pair position of the variant according to human reference sequence (GRCh37), Hg19; EA: effect allele; EAF: effect allele frequency; MAF: minor allele frequency; β: per allele effect in standard deviations of standardized log-transformed 25OHD or 1,25 dihydroxyvitamin D; P: strength of evidence against the null hypothesis of no associations between variant and 25OHD (ie P-value) from standard linear regression; β.J: per allele effect estimated from joint analysis of conditionally associated snps; P.J: Strength of evidence against the null hypothesis of no association between the variant and 25OHD as estimated by conditional and joint genome-wide association analysis (i.e. P-value);VAR.J: Proportion of variance explained by the conditionally associated variant; A.GENE: The name of the gene situated closest to variant that has smallest P-value of all conditionally independent variants present in the same locus; FUNCTION: Functional annotation of the conditionally independent variant.